University of Southern Denmark
IMADA
Arthur Zimek
Jonatan Møller Gøttcke, Jonas Herskind Sejr

DM566/DM868/DM870: Data Mining and Machine Learning

Spring term 2019

Exercise 15: Information gain, Perceptron, SVMs, Practical

Exercise 15-1 Information gain

In this exercise, we want to look more closely at the information gain measure.

Let T be a set of n training objects with the attributes $A_1, \ldots, A_a$ and the k classes c_1 to c_k.

Let $\{T_i^A \mid i \in \{1, \ldots, m_A\}\}$ be the disjoint, complete partitioning of T produced by a split on attribute A (where m_A is the number of disjoint values of A).

(a) *Uniform distribution*
Compute *entropy*(T), *entropy*(T_i^A) for $i \in \{1 \ldots m_A\}$ as well as *information-gain*(T, A) given the assumption that the class membership of T is uniformly distributed and independent of the values of A. Interpret your result!

(b) *Additional uniform distribution*
We want to analyze how the number of different values influences the information gain. For this, we compare two attributes, attribute A with m_A values and attribute A' with $m_{A'} = m_A + 1$ values, where the relative frequencies in A' in values 1 to m_A are identical to that of A and in the additional value $m_{A'}$ there is a uniform distribution of the classes.
How does *information-gain*(T, A) differ from *information-gain*(T, A')? Interpret your result!

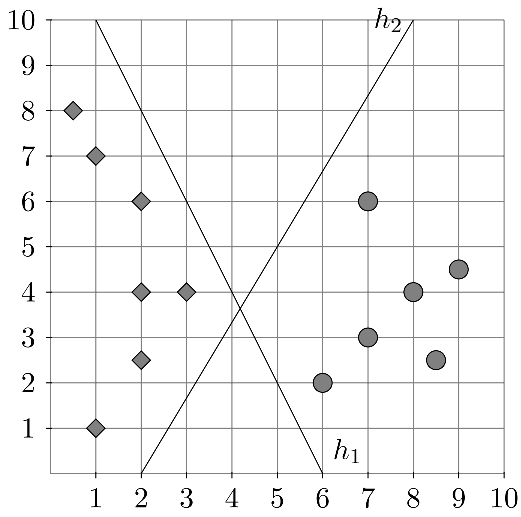
(c) *Attributes with many values*
Let A be an attribute with random values, not correlated to the class of the objects. Furthermore, let A have enough values, such than not any two instances of the training set share the same value of A. What happens in this situation when building the decision tree? What is problematic with this situation?

Exercise 15-2 Perceptron

Sketch two trained perceptrons (no hidden layer!) that can represent for two Boolean variables $x_1, x_2 \in \{0, 1\}$ the AND and the OR function ($x_1 \wedge x_2$ bzw. $x_1 \vee x_2$), respectively.

Exercise 15-3 Support vectors and margin

Consider the following dataset with points from two classes c_1 (diamonds) and c_2 (circles).



(a) Give the equations for hyperplanes h_1 and h_2.

(b) Name all the support vectors for h_1 and h_2.

(c) Which of the two hyperplanes is better at separating the two classes based on the margin?

(d) Find the best separating hyperplane for this dataset, give its equation, and show the corresponding support vectors.

Hint: You can do this without solving the optimization problem by considering the convex hull of each class and the possible hyperplanes at the boundary of the two classes.

Exercise 15-4 Neural networks and support vector machines – Practical

Bring your laptop for this interactive lab session in the exercise class. Work with some toolbox for classification (e.g., R, Python, WEKA) to study the impact of different settings on the behavior of neural networks and support vector machines on some dataset.

(a) Choose datasets of considerably different size: what do you observe for the training time?

(b) Choose different setup for the classifiers (different kernels/different number of layers):

- What do you observe for the training time?
- What do you observe for the apparent error and the true error?